

United States Patent Application for

**METHOD AND APPARATUS FOR EFFICIENT
TRANSMISSION OF RESULT SET DATA**

Inventors

Sreenivas Gollapudi

Jonathan Klein

Chi Ku

Assignee

Oracle International Corporation
500 Oracle Parkway
Mail Stop 50p7
Redwood Shores, CA 94065

METHOD AND APPARATUS FOR EFFICIENT TRANSMISSION OF RESULT SET DATA

5

Inventors: Sreenivas Gollapudi, Jonathan Klein & Chi Ku

BACKGROUND

This invention relates generally to the field of data communications. More
10 particularly, a method and apparatus are provided for transporting data from a
data server.

A data server (e.g., a database, an origin server) is configured to serve data
to clients and other servers. For example, a data server may receive a query from
a client (e.g., a SELECT statement), retrieve data that satisfy the query from disk
15 or other storage, and serve the data to the client. The data may comprise any
number of rows, or records, and each row may contain any number of columns, or
attributes.

Traditionally, such data (e.g., a row set, a result set) is shipped from the
data server one row at a time. However, before transmission, each row or column
20 is usually decomposed into its constituent columns, which may be processed to
convert data types, apply business logic, retrieve related data, format a column's
data, etc. For wide rows (e.g., rows having tens or hundreds of columns), a
significant amount of time may be expended performing the processing on the
server prior to its transmission. And, in some cases, the client may be configured
25 or configurable to perform the same processing.

At the client or other recipient, data are received one row at a time, and
may be broken down into individual columns and consumed as needed for the

controlling application or utility. The data may thus be disassembled twice – once at the server and once at the client – before being consumed. For simple data (e.g., those requiring little or no processing on the data server), two disassembly operations may be superfluous.

5

SUMMARY

In one embodiment of the invention, a method and apparatus are provided for efficiently transmitting a result set. A data server receives a data request from a client or another server. The data server determines whether a threshold portion
10 of the contents of each requested row or record is being retrieved (e.g., X% of the number of columns in a table or X% of the total amount of data in a table row). The requested rows are retrieved in the same format in which they are stored (e.g., within a disk block), and streamed to the client, one row at a time. The client is then responsible for disassembling a row into its columns, reordering them if
15 necessary, and performing any necessary post-processing (e.g., datatype conversion, data formatting) before consuming the data.

The server may inform the data requestor of the storage format of the data each time a data request is received, when a session is established between the requestor and the server, or at some other time.

20 If a result set requires post-processing that is too complex for the client or, possibly, if the amount of data requested is less than the applicable threshold, the server may disassemble the requested rows, reorder them as necessary and perform the post-processing before transmitting the result set to the client.

25

DESCRIPTION OF THE FIGURES

FIG. 1 depicts a computing environment in which an embodiment of the present invention may be implemented.

FIG. 2 is a flowchart illustrating one method of efficiently transmitting a result set from a data server to a client, in accordance with an embodiment of the invention.

5

DETAILED DESCRIPTION

The following description is presented to enable any person skilled in the art to make and use the invention, and is provided in the context of particular applications of the invention and their requirements. Various modifications to the disclosed embodiments will be readily apparent to those skilled in the art and the
10 general principles defined herein may be applied to other embodiments and applications without departing from the scope of the present invention. Thus, the present invention is not intended to be limited to the embodiments shown, but is to be accorded the widest scope consistent with the principles and features disclosed herein.

15 The program environment in which a present embodiment of the invention is executed illustratively incorporates a data server, a general-purpose computer or a special purpose device such as a hand-held computer. Details of such devices (e.g., processor, memory, data storage, display, communication interface) may be omitted for the sake of clarity.

20 It should also be understood that the techniques of the present invention may be implemented using a variety of technologies. For example, the methods described herein may be implemented in software executing on a computer system, or implemented in hardware utilizing either a combination of microprocessors or other specially designed application specific integrated
25 circuits, programmable logic devices, or various combinations thereof. In particular, the methods described herein may be implemented by a series of computer-executable instructions residing on a suitable computer-readable medium. Suitable computer-readable media may include volatile (e.g., RAM)

and/or non-volatile (e.g., ROM, disk) memory, carrier waves and transmission media (e.g., copper wire, coaxial cable, fiber optic media). Exemplary carrier waves may take the form of electrical, electromagnetic or optical signals conveying digital data streams along a local network, a publicly accessible
5 network such as the Internet or some other communication link.

In one embodiment of the invention, a method and apparatus are provided for transporting data (e.g., a row set, a result set) more efficiently from a data server or other source. Illustratively, the data to be transported may reside in a database such as an Oracle Database by Oracle Corporation.

10 In this embodiment, data are shipped from the server to a client or other recipient, such as another server, in substantially the same form in which it was stored on disk or other storage device. For example, the data may be transmitted one row at a time, but will be transmitted as retrieved from storage, without being broken into columns and subjected to per-column processing.

15 So that the client can unmarshal the data, the server informs the client of the configuration of the disk block, or other format, in which the data are stored. Upon receipt of the data, the client is able to extract the data and process it as needed. Because processing of the data is moved from the server, a significant decrease can be achieved in the number of instructions that must be executed on
20 the server in order to serve the data.

To facilitate implementation of the invention, the client may be configured with an interface, such as Oracle Call Interface (OCI), that is tightly coupled to the database or data source maintained on the server. Many interfaces, such as Java Database Connectivity (JDBC), Open Database Connectivity (ODBC) and
25 Command Line Interface (CLI) are more generic client interfaces, but may be modified to include the functionality described herein.

FIG. 1 depicts an environment in which an illustrative embodiment of the invention may be implemented. In this embodiment, server 110 is coupled to

client 120 by a communication link, which may be wired or wireless and may comprise a network, point-to-point link or other connection.

Server 110 is a database server, origin server, web server or other computer server configured to serve data in response to requests from any sources
 5 (e.g., clients, other servers). Client 110 is a computer system (e.g., handheld, desktop, mobile) or other device equipped with a processor and software or firmware for requesting and processing data from server 110.

Server 110 includes server interface 112 for receiving data requests and sending result sets in response to requests. Server 110 also includes one or more
 10 databases, datasets or other data compilations 118 comprising data to be served or transmitted to clients, servers and/or other entities.

Client 120 includes client interface 122, which is configured to communication with server interface 112 to request and receive data from server 110. Client interface 122 includes column (or field) processor or processing
 15 module 124, which is configured to apply per-column processing to some or all columns or fields of database rows or records received from the server. In this embodiment of the invention, server 110 may perform per-column processing in some cases; therefore, server interface 112 may include column processor 114. The client interface may be OCI (Oracle Call Interface), by Oracle Corporation.

20 Data requests may be submitted to server 110 from client 120, and other sources, in any format understood by server interface 112 (e.g., as SQL statements, queries). As described herein, a result set for a request may be returned in substantially the same format as it is received or retrieved from database 118.

25 In an embodiment of the invention, the selectivity of a query or request submitted to the data server may be defined as the ratio between the number of columns in the select list and the number of columns in the database table to which the query is applied. Or, the selectivity may be defined as the ratio

between the amount of data in the columns of the select list (e.g., cumulative size of all selected columns) and the total amount of data in a row of the table (e.g., total size of all columns).

If the selectivity of the query is relatively low (e.g., less than 50%, less than 80%), the result set of the query may be processed traditionally – with disassembly of rows and per-column processing being performed on the server. If the selectivity is relatively high, then the result set may be sent to the client as described herein – without being disassembled on the server, without being subjected to per-column processing on the server and in basically the same format in which the data were stored. In another embodiment of the invention, most or all result sets may be transmitted as described herein.

In an embodiment of the invention, when the server transmits a result set in the form in which it was stored, without being disassembled and processed on the server, entire rows or records are sent. In other words, all columns or attributes in each selected row are sent, even if only a subset of all columns were selected.

Another embodiment of the invention is applied only when a data request targets “real” columns in a data table, not pseudo-columns. A real column is a column that is stored as part of the table. A pseudo-column is generated by applying a function or other processing to one or more real columns. Pseudo-columns are not stored with the table. As one exception, however, data requests may employ or identify data by row id, a special pseudo-column that does not need to be computed from a row’s real columns.

To enable implementation of an embodiment of the invention, clients, servers and other entities that request data from a server are informed of the disk row format, or other storage format, of data stored on the server. For example, rows in a database table may be stored as a contiguous set of the row’s column

values. For each row of the table, the columns are generally stored in the same order.

When a data request is received, one or more responsive rows are identified and retrieved. Instead of reordering the rows as specified in the request, 5 however, the rows are transmitted in the order and form in which they are retrieved from storage. The requestor is responsible for disassembling each row as it is received, reordering the columns as needed and performing any post-processing (e.g., to convert a data type, to apply a desired presentation format).

FIG. 2 demonstrates a method of efficiently transmitting data from a data 10 server, according to one embodiment of the invention. In this embodiment, a data request is received as a query from a client, but the embodiment is compatible with multiple data request types and formats and can receive them from other servers as well as clients.

Applying this embodiment of the invention reduces the code path in the 15 server and reduces the server's overall response time to a data request.

In state 202, a client establishes a session with a data server that stores data needed by the client. The server may employ any type or types of storage devices, such as disk, tape, etc. As part of the process of establishing a session, the server informs the client of the row format for the database table(s) the client 20 will access on the server. Illustratively, the format of a table may be sent as a sequence of tuples having the form *<size, data>*, where *size* is the size of a column (or attribute or field) and *data* is the datatype of the column.

In one alternative embodiment of the invention, the server notifies a client of the row format of a database table each time (or the first time) the client 25 submits a query involving the table. The timing or manner in which the client is informed of the row format is not limited in any embodiment of the invention.

During state 202, the client may also inform the server of its ability to process data retrieved from the server (e.g., to convert a data type, apply business logic).

5 In state 204, the server (e.g., a server interface) receives a data request from the client. Illustratively, the request may comprise a SELECT statement identifying one or more rows of a database table and one or more columns within those rows.

In state 206, the server determines whether the amount of columnar data selected for the table meets a threshold, based on the number of columns or the
10 size of the columns. An illustrative threshold is 80%. If the selected amount of data is less than the threshold, the method advances to state 220.

In state 208, the server attempts to determine a level of complexity needed for post-processing of the requested data. Post-processing comprises various processing that may be applied to retrieved data to convert a data type, format a
15 column's data, retrieve related data from another table, apply some business logic, etc.

Different clients or data requestors may be capable of different levels of processing. If the level of complexity required for post-processing exceeds what the server knows or believes the present client is capable of, then the post-
20 processing may be done on the server, not the client. In this case, the illustrated method advances to state 220. Otherwise, the client is deemed capable of performing any necessary post-processing, and the method continues with state 210.

In state 210, the server retrieves the requested rows from storage (e.g.,
25 cache, disk), in the format (e.g., column order) in which they are stored. Thus, a disk block may be retrieved and one or more rows extracted from it. The columns of a row are not reordered to match the client's request format. The server simply identifies each row that meets the data request, and retrieves the entire row.

In state 212, the server transmits each row to the client. Illustratively, each row is transmitted as it was retrieved from storage (i.e., with the columns in the same order in which they were stored), and without disassembling or performing post-processing. In an alternative embodiment of the invention, the
5 server may only retrieve or send those columns that are requested, rather than sending the entire contents of each row.

The client receives one row at a time and disassembles them as necessary to retrieve the columns. Unneeded columns may be discarded, and the others are reordered and post-processed as necessary before being consumed. The method
10 then ends.

In state 220, the server retrieves the requested data, disassembles rows to retrieve, reorder and post-process columns, then transmits the requested columns of each requested row, in the order requested by the client. The method then ends.

15 The foregoing descriptions of embodiments of the invention have been presented for purposes of illustration and description only. They are not intended to be exhaustive or to limit the invention to the forms disclosed. Accordingly, the above disclosure is not intended to limit the invention; the scope of the invention is defined by the appended claims.

20 For example, in a tightly coupled or clustered environment, an embodiment of the invention may be applied as a means for moving data between nodes. Thus, if a first node is relatively busy compared to a second node, data may be transmitted from the first node to the second node as described above, so that post-processing of the data can be performed on the second node instead of
25 the first.